

# UC Davis

## UC Davis Previously Published Works

### Title

Complete genome sequence of Syntrophothermus lipocalidus type strain (TGB-C1).

### Permalink

<https://escholarship.org/uc/item/6h65c1r2>

### Journal

Standards in genomic sciences, 3(3)

### ISSN

1944-3277

### Authors

Djao, Olivier Duplex Ngatchou  
Zhang, Xiaojing  
Lucas, Susan  
et al.

### Publication Date

2010-12-01

### DOI

10.4056/sigs.1233249

Peer reviewed

## Complete genome sequence of *Syntrophothermus lipocalidus* type strain (TGB-C1<sup>T</sup>)

Olivier Duplex Ngatchou Djao<sup>1</sup>, Xiaojing Zhang<sup>2</sup>, Susan Lucas<sup>3</sup>, Alla Lapidus<sup>3</sup>, Tijana Glavina Del Rio<sup>3</sup>, Matt Nolan<sup>3</sup>, Hope Tice<sup>3</sup>, Jan-Fang Cheng<sup>3</sup>, Cliff Han<sup>2</sup>, Roxanne Tapia<sup>2</sup>, Lynne Goodwin<sup>2,3</sup>, Sam Pitluck<sup>3</sup>, Konstantinos Liolios<sup>3</sup>, Natalia Ivanova<sup>3</sup>, Konstantinos Mavromatis<sup>3</sup>, Natalia Mikhailova<sup>3</sup>, Galina Ovchinnikova<sup>3</sup>, Amrita Pati<sup>3</sup>, Evelyne Brambilla<sup>4</sup>, Amy Chen<sup>5</sup>, Krishna Palaniappan<sup>5</sup>, Miriam Land<sup>3,6</sup>, Loren Hauser<sup>3,6</sup>, Yun-Juan Chang<sup>3,5</sup>, Cynthia D. Jeffries<sup>3,6</sup>, Manfred Rohde<sup>1</sup>, Johannes Sikorski<sup>4</sup>, Stefan Spring<sup>4</sup>, Markus Göker<sup>4</sup>, John C. Detter<sup>3</sup>, Tanja Woyke<sup>3</sup>, James Bristow<sup>2</sup>, Jonathan A. Eisen<sup>2,7</sup>, Victor Markowitz<sup>4</sup>, Philip Hugenholtz<sup>4</sup>, Nikos C. Kyrpides<sup>3</sup>, and Hans-Peter Klenk<sup>4\*</sup>

<sup>1</sup> HZI – Helmholtz Centre for Infection Research, Braunschweig, Germany

<sup>2</sup> Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA

<sup>3</sup> DOE Joint Genome Institute, Walnut Creek, California, USA

<sup>4</sup> DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany

<sup>5</sup> Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA

<sup>6</sup> Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

<sup>7</sup> University of California Davis Genome Center, Davis, California, USA

\*Corresponding author: Hans-Peter Klenk

**Keywords:** anaerobic, motile, Gram-negative, syntrophism with methanogen, crotonate, butyrate, isobutyrate, *Syntrophomonadaceae*, GEBA

*Syntrophothermus lipocalidus* Sekiguchi *et al.* 2000 is the type species of the genus *Syntrophothermus*. The species is of interest because of its strictly anaerobic lifestyle, its participation in the primary step of the degradation of organic matters, and for releasing products which serve as substrates for other microorganisms. It also contributes significantly to maintain a regular pH in its environment by removing the fatty acids through  $\beta$ -oxidation. The strain is able to metabolize isobutyrate and butyrate, which are the substrate and the product of degradation of the substrate, respectively. This is the first complete genome sequence of a member of the genus *Syntrophothermus* and the second in the family *Syntrophomonadaceae*. Here we describe the features of this organism, together with the complete genome sequence and annotation. The 2,405,559 bp long genome with its 2,385 protein-coding and 55 RNA genes is a part of the *Genomic Encyclopedia of Bacteria and Archaea* project.

### Introduction

Strain TGB-C1<sup>T</sup> (= DSM 12680) is the type strain of *Syntrophothermus lipocalidus* [1] which in turn is the type species of the genus *Syntrophothermus* [2]. Currently, this is the only species placed in the genus *Syntrophothermus*. The genus name derives from the Greek words “syn”, together with, “trophos”, one who feeds, and “thermus”, hot, referring to a thermophilic bacterium growing in syntrophic association with hydrogenotrophic organisms at high temperature of around 55°C [1]. The species epithet derives from the Greek word “lipos”, fat, and from the Latin adjective “calidus”, expert, re-

ferring to the organisms trait of specifically utilizing fatty acids [1]. Strain TGB-C1<sup>T</sup> was isolated from granular sludge in a thermophilic upflow anaerobic sludge blanket (UASB) [1]. No further cultivated strains belonging to the species *S. lipocalidus* have been described so far. Here we describe the features of this organism, together with the complete genome sequence and annotation.

### Classification and features

The 16S rRNA gene sequence of strain TGB-C1<sup>T</sup> revealed an only distant relationship with the other

representatives of the family *Syntrophomonadaceae* [1] (Figure 1), with *Thermosyntropha lipolytica* [10] showing the highest degree of sequence similarity (88.1%). The sequence distances of strain TGB-C1<sup>T</sup> to other members of this family were 13.6% with *Syntrophomonas wolfei* subsp. *wolfei*, 14.0% with *S. bryantii*, and 14.8% with *S. sapovorans*, respectively [1]. Further analysis showed 98% 16S rRNA gene sequence identity with an uncultured bacterium represented by clone AR80B63 (AB539943) from the high-temperature Yabase oil field in Japan. The sequence of the 16S rRNA gene of strain TGB-C1<sup>T</sup> is identical with two unclassified sequences from an hydrothermal vent metagenome LCHCB.C3615 [11] and from human gut metagenome DNA (contig sequence: F2-Y\_011332) [12] (status August 2010), indicating that members of the species, genus and even family are widely represented in the habitats screened so far.

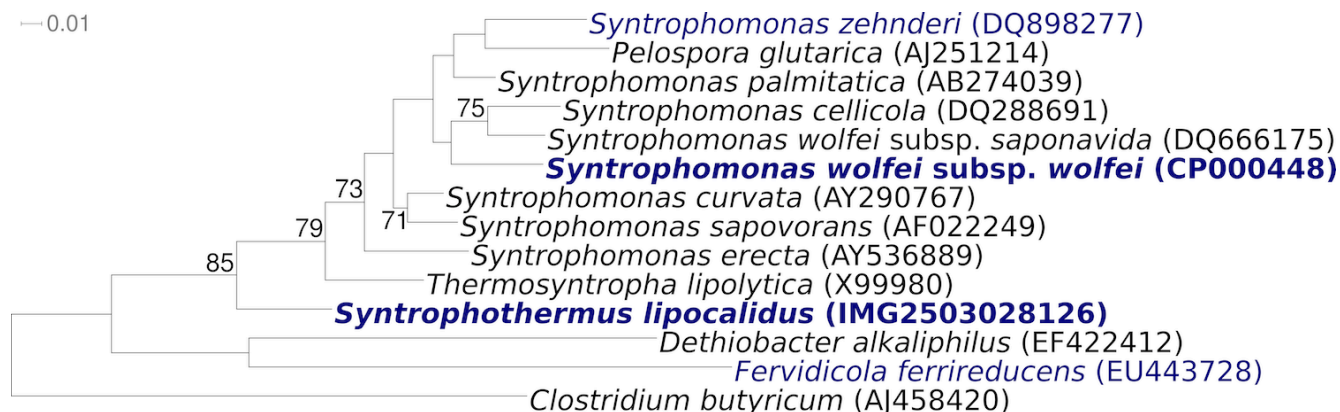
A representative genomic 16S rRNA sequence of *S. lipocalidus* TGB-C1<sup>T</sup> was also compared using BLAST with the most recent release of the GreenGenes database [13] and the relative frequencies of taxa and keywords, weighted by BLAST scores, were determined. The five most frequent genera were *Moorella* (44.1%), *Syntrophomonas* (33.8%), *Clostridium* (6.0%), *Syntrophothermus* (5.6%) and *Carboxydocella* (3.5%). The species yielding the highest score was *Moorella thermoautotrophica*. The five most frequent keywords within the labels of environmental samples which yielded hits were 'microbial' (5.5%), 'anaerobic' (4.2%), 'rice' (2.9%), 'soil' (2.8%) and 'populations' (2.8%). The three most frequent keywords within the labels of environmental samples which yielded hits of a higher score than the highest scoring species were 'temperature' (8.2%), 'acetate, coupled, evidence, field, hydrogenotrophic, methanogenesis, oil, oxidation, petroleum, reservoir, syntrophic, yabase' (5.0%) and 'dependent, hot, muddy, reducing, sediment, southwestern, spring, succession, sulfate, taiwan' (3.2%). These keywords largely fit to what is known about the ecology and physiology of strain TGB-C1<sup>T</sup> [1].

Figure 1 shows the phylogenetic neighborhood of *S. lipocalidus* TGB-C1<sup>T</sup>, in a 16S rRNA based tree. The sequences of the two 16S rRNA gene copies in the genome differ from each other by up to two nucleotides, and differ by up to two nucleotides from the previously published 16S rRNA sequence (AB021305).

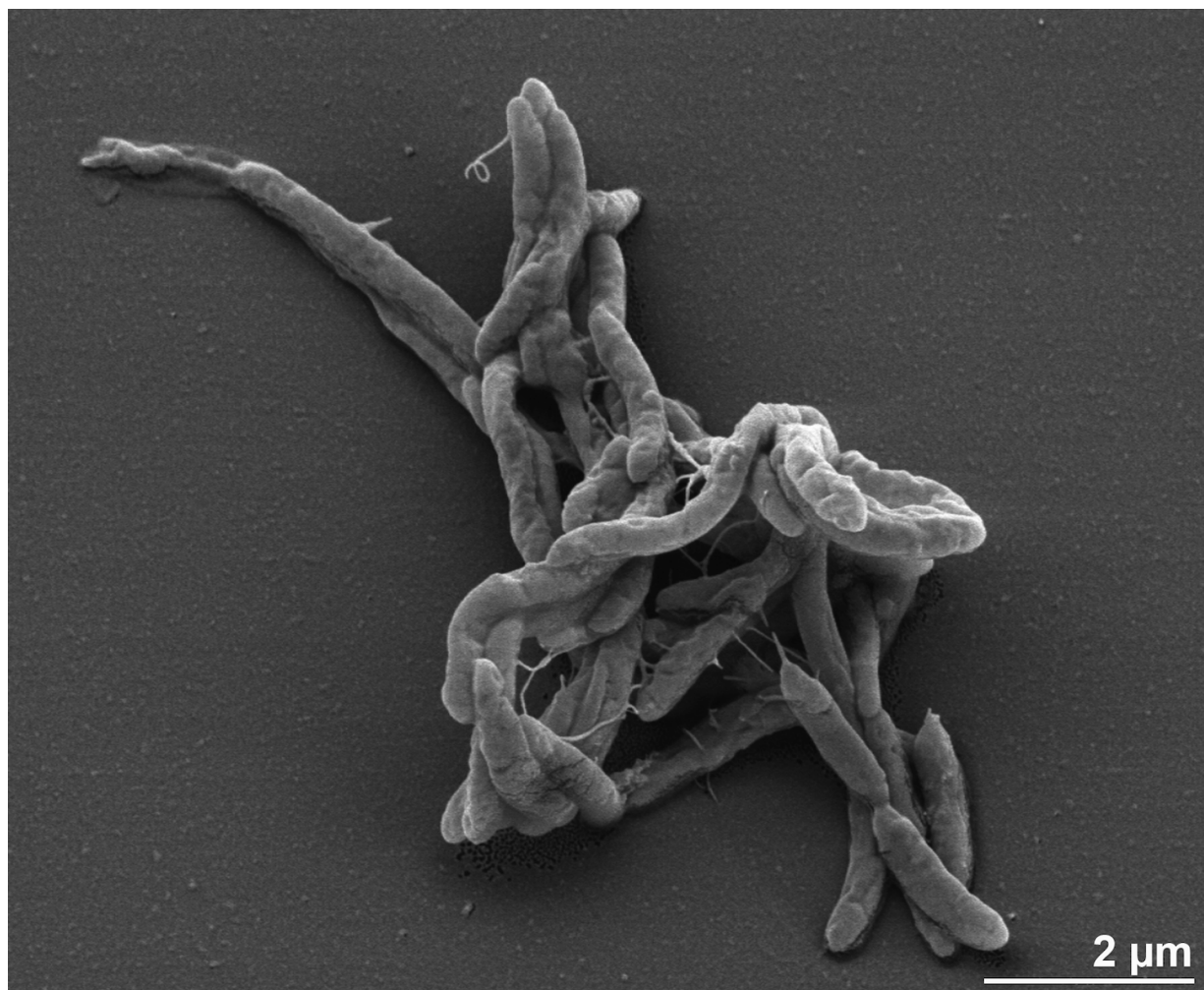
Cells of strain TGB-C1<sup>T</sup> are Gram-negative, slightly curved rods with round ends and weakly motile with flagella, 2.4 - 4.0 µm long and 0.4 - 0.5 µm wide (Figure 2 and Table 1) [1], occurring singly or in pairs. Roll-tube isolation revealed the presence of small white colonies, lens-shaped and 0.1 - 0.2 mm in diameter [1]. The growth rate of the strain TGB-C1<sup>T</sup> on 10 mM crotonate was  $0.93 \pm 0.01 \text{ d}^{-1}$ . Strain TGB-C1<sup>T</sup> is strictly anaerobic [1]. It grows on crotonate at temperatures between 45°C and 60°C, with the optimum at 55°C. The pH<sub>25°C</sub> range for growth is 5.8-7.5, with an optimum at 6.5-7.0 [1]. Strain TGB-C1<sup>T</sup> metabolizes in two ways, in pure culture only in the presence of the unsaturated fatty acid crotonate and in co-culture with *Methanobacterium thermoautotrophicum* strain ΔH in the presence of saturated fatty acids [1]. In pure culture, the fermentation products are acetate and butyrate in equimolar amounts. In co-culture with *M. thermoautotrophicum*, the substrates used are butyrate, straight-chain fatty acids from C<sub>4</sub> to C<sub>10</sub> and isobutyrate [1]. By oxidizing fatty acids, *S. lipocalidus* produces acetate and hydrogen [1], the latter of which is then scavenged by the syntrophic methanogen *M. thermoautotrophicum* [1]. Syntrophic hydrogenotrophic interactions with bacteria from the genus *Methanobacterium* have been also observed in the genome sequenced bacterium *Aminobacterium colombiense* strain ALA-1<sup>T</sup> from the phylum *Synergistetes* [26]. *S. lipocalidus* is the only species in the family *Syntrophomonadaceae* that is able to metabolize isobutyrate [2]. Neither yeast extract nor tryptone significantly stimulates growth [1]. In the presence of butyrate as electron donor, the following compounds do not serve as electron acceptors: sulfate, nitrate, sulfite, thiosulfate, fumarate, Fe(III)-nitritotriacetate [1]. Cell growth is inhibited by ampicillin, chloramphenicol, kanamycin, neomycin, rifampin or vancomycin (each 50 µg ml<sup>-1</sup>) [1].

### Chemotaxonomy

To date, no experimental reports have specified the lipid composition of the cell envelope of strain TGB-C1<sup>T</sup>. Nevertheless, the cell envelope of the strain TGB-C1<sup>T</sup> was Gram-negative stained, although electron micrographs and the 16S rRNA analysis showed that the strain was affiliated to the Gram-positive bacteria [1]. This feature was also observed for another member of the family *Syntrophomonadaceae*, *S. bryantii* [22,27]. The cell envelope is composed of the cytoplasmic membrane, an electron-dense layer, which is most probably made of peptidoglycan, and an electron-dense outermost wall [1].



**Figure 1.** Phylogenetic tree highlighting the position of *S. lipocalidus* TGB-C1<sup>T</sup> relative to the type strains within the family *Syntrophomonadaceae*. The trees were inferred from 1,434 aligned characters [3,4] of the 16S rRNA gene sequence under the maximum likelihood criterion [5] and rooted in accordance with the current taxonomy [6]. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches are support values from 1,000 bootstrap replicates [7] if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [8] are shown in blue, published genomes in bold [9].



**Figure 2.** Scanning electron micrograph of *S. lipocalidus* TGB-C1<sup>T</sup>

**Table 1.** Classification and general features of *S. lipocalidus* TGB-C1 in according with the MIGS recommendations [14]

MIGS ID	Property	Term	Evidence code
		Domain <i>Bacteria</i>	TAS [15]
		Phylum <i>Firmicutes</i>	TAS [16,17]
		Class <i>Clostridia</i>	TAS [18,19]
	Current classification	Order <i>Clostridiales</i>	TAS [20,21]
		Family <i>Syntrophomonadaceae</i>	TAS [22,23]
		Genus <i>Syntrophothermus</i>	TAS [1]
		Species <i>Syntrophothermus lipocalidus</i>	TAS [1]
		Type strain TGB-C1	TAS [1]
	Gram stain	negative	TAS [1]
	Cell shape	slightly curved rods with round ends	TAS [1]
	Motility	weakly motile by flagella	TAS [1]
	Sporulation	none	TAS [1]
	Temperature range	45°C–60°C	TAS [1]
	Optimum temperature	55°C	TAS [1]
	Salinity	< 0.5% NaCl	TAS [1]
MIGS-22	Oxygen requirement	obligately anaerobic	TAS [1]
	Carbon source	crotonate in pure culture; fatty acids with 4-10 carbon atoms including isobutyrate in syntrophy	TAS [1]
	Energy source	crotonate	TAS [1]
MIGS-6	Habitat	not reported	NAS
MIGS-15	Biotic relationship	syntrophic with methanogens	NAS
MIGS-14	Pathogenicity	not reported	NAS
	Biosafety level	1	TAS [24]
	Isolation	granular sludge in a thermophilic upflow anaerobic sludge blanket (UASB) reactor	TAS [1]
MIGS-4	Geographic location	most probably Japan	TAS [1]
MIGS-5	Sample collection time	2000 or before	TAS [1]
MIGS-4.1	Latitude	not reported	
MIGS-4.2	Longitude		NAS
MIGS-4.3	Depth	not reported	NAS
MIGS-4.4	Altitude	not reported	NAS

Evidence codes - IDA: Inferred from Direct Assay (first time in publication); TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from of the Gene Ontology project [25]. If the evidence code is IDA, then the property was directly observed by one of the authors or an expert mentioned in the acknowledgements.

## Genome sequencing and annotation

### Genome project history

This organism was selected for sequencing on the basis of its phylogenetic position [28], and is part of the *Genomic Encyclopedia of Bacteria and Archaea* project [29]. The genome project is deposited in the Genome OnLine Database [8] and the com-

plete genome sequence is deposited in GenBank. Sequencing, finishing and annotation were performed by the DOE Joint Genome Institute (JGI). A summary of the project information is shown in Table 2.



**Table 2.** Genome sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Finished
MIGS-28	Libraries used	Three genomic libraries: 454 pyrosequence standard library and; paired end library (10.2 kb insert size); Illumina standard library
MIGS-29	Sequencing platforms	454 GS FLX Titanium, Illumina GAii
MIGS-31.2	Sequencing coverage	103.3× pyrosequence, 81.3× Illumina
MIGS-30	Assemblers	Newbler version 2.1-PreRelease-4-28-2009, Velvet, phrap
MIGS-32	Gene calling method	Prodigal 1.4, GenePRIMP
	INSDC ID	CP002048
	Genbank Date of Release	June 7, 2010
	GOLD ID	Gc012392
	NCBI project ID	37873
	Database: IMG-GEBA	2502957035
MIGS-13	Source material identifier	DSM 12680
	Project relevance	Tree of Life, GEBA

### Growth conditions and DNA isolation

*S. lipocalidus* TGB-C1<sup>T</sup>, DSM 12680, was grown anaerobically in DSMZ medium 870 (*Syntropho-thermus* medium) [30] at 55°C. DNA was isolated from 0.5-1 g of cell paste using the Jetflex Genomic DNA Purification kit (GENOMED 600100) following the standard protocol as recommended by the manufacturer, with 30 min incubation at 58°C for cell lysis.

### Genome sequencing and assembly

The genome was sequenced using a combination of Illumina and 454 sequencing platforms. All general aspects of library construction and sequencing can be found at the JGI website [31]. Pyrosequencing reads were assembled using the Newbler assembler version 2.1-PreRelease-4-28-2009-gcc-3.4.6-threads (Roche). The initial Newbler assembly consisting of 16 contigs in one scaffold was converted into a phrap assembly by making fake reads from the consensus, collecting the read pairs in the 454 paired end library. Illumina GAii sequencing data (704 Mb) was assembled with Velvet [32] and the consensus sequences were shredded into 1.5 kb overlapped fake reads and assembled together with the 454 data. 454 draft assembly was based on 248.9 Mb 454 draft data and all of the 454 paired end data. Newbler parameters are -consed -a 50 -l 350 -g -m -ml 20. The Phred/Phrap/Consed software package [33] was used for sequence assembly and quality assessment in the following finishing process. After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC). Possible mis-assemblies were corrected with gapResolution [31], Dupfinisher, or sequencing cloned bridging PCR fragments with subcloning or transposon bombing (Epicentre Biotechnologies, Madison, WI)

[34]. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR primer walks (J.-F.Chang, unpublished). A total of 37 additional reactions were necessary to close gaps and to raise the quality of the finished sequence. Illumina reads were also used to correct potential base errors and increase consensus quality using a software Polisher developed at JGI [35]. The error rate of the completed genome sequence is less than 1 in 100,000. Together, the combination of the Illumina and 454 sequencing platforms provided 184.6 × coverage of the genome. Final assembly contains 815,143 pyrosequence and 5,434,428 Illumina reads.

### Genome annotation

Genes were identified using Prodigal [36] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI [37]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction analysis and functional annotation was performed within the (IMG-ER) platform [38].

### Genome properties

The genome consists of a 2,405,559 bp long chromosome with a 51.0% GC content (Table 3 and Figure 3). Of the 2,440 genes predicted, 2,385 were protein-coding genes, and 55 RNAs; 72 pseudogenes were also identified. The majority of the protein-coding genes (70.7%) were assigned with a putative function while the remaining ones were annotated as hypothetical proteins. The distribution of genes into COGs functional categories is presented in Table 4.

Table 3. Genome Statistics

Attribute	Value	% of Total
Genome size (bp)	2,405,559	100.00%
DNA coding region (bp)	2,078,709	86.41%
DNA G+C content (bp)	1,226,580	50.99%
Number of replicons	1	
Extrachromosomal elements	0	
Total genes	2,440	100.00%
RNA genes	55	2.25%
rRNA operons	2	
Protein-coding genes	2,385	97.75%
Pseudo genes	72	2.95%
Genes with function prediction	1,726	70.74%
Genes in paralog clusters	348	14.26%
Genes assigned to COGs	1,767	72.42%
Genes assigned Pfam domains	1,912	78.26%
Genes with signal peptides	603	24.71%
Genes with transmembrane helices	545	22.34%
CRISPR repeats	2	

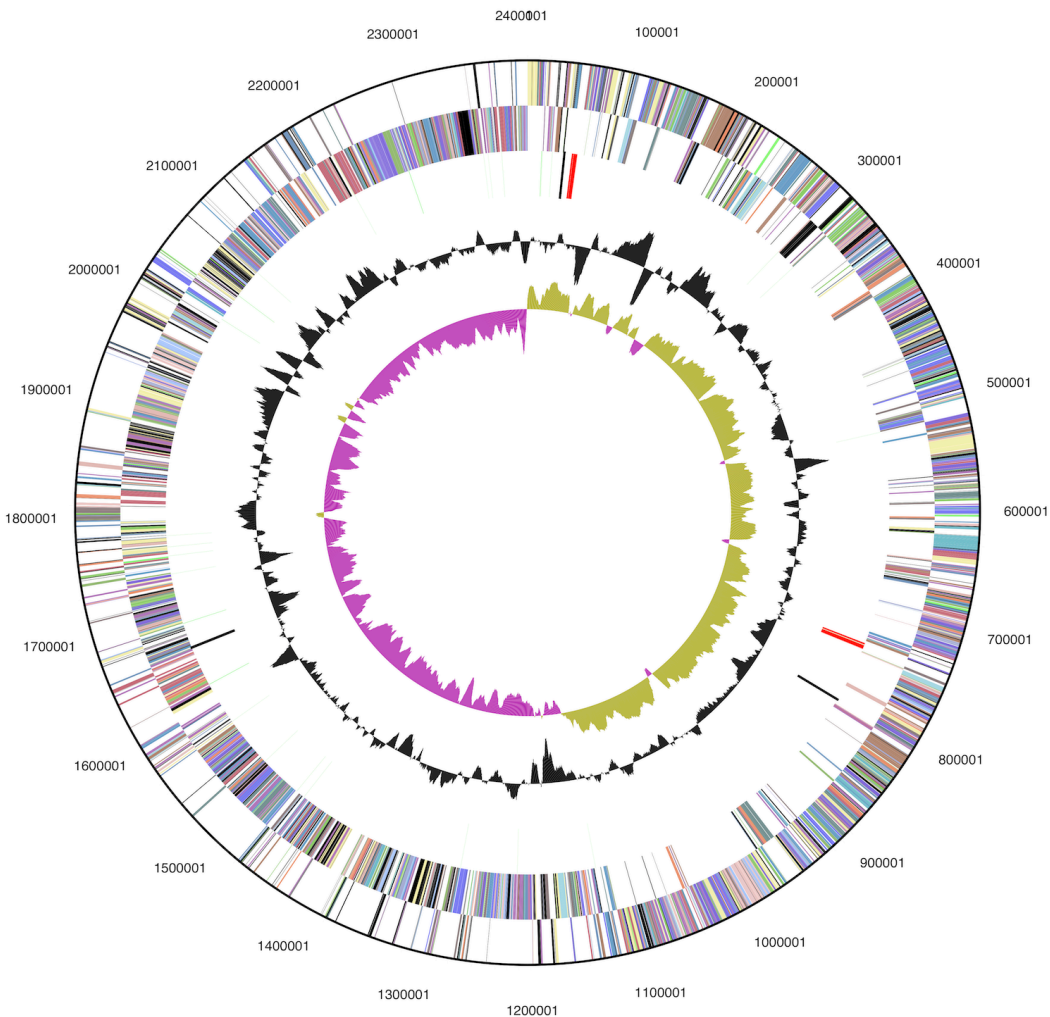


Figure 3. Graphical circular map of the genome. From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew.

**Table 4.** Number of genes associated with the general COG functional categories

Code	value	%age	Description
J	144	7.4	Translation, ribosomal structure and biogenesis
A	0	0.0	RNA processing and modification
K	113	5.8	Transcription
L	123	6.3	Replication, recombination and repair
B	4	0.2	Chromatin structure and dynamics
D	32	1.6	Cell cycle control, cell division, chromosome partitioning
Y	0	0.0	Nuclear structure
V	33	1.7	Defense mechanisms
T	107	5.5	Signal transduction mechanisms
M	96	4.9	Cell wall/membrane/envelope biogenesis
N	81	4.1	Cell motility
Z	0	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	66	3.4	Intracellular trafficking and secretion, and vesicular transport
O	74	3.8	Posttranslational modification, protein turnover, chaperones
C	144	7.4	Energy production and conversion
G	67	3.4	Carbohydrate transport and metabolism
E	144	7.4	Amino acid transport and metabolism
F	58	3.0	Nucleotide transport and metabolism
H	112	5.7	Coenzyme transport and metabolism
I	98	5.0	Lipid transport and metabolism
P	70	3.6	Inorganic ion transport and metabolism
Q	26	1.3	Secondary metabolites biosynthesis, transport and catabolism
R	205	10.5	General function prediction only
S	158	8.1	Function unknown
-	673	27.6	Not in COGs

## Acknowledgements

We would like to gratefully acknowledge the help of Maren Schröder (DSMZ) in cultivation of the strain. This work was performed under the auspices of the US Department of Energy Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence

Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396, UT-Battelle, and Oak Ridge National Laboratory under contract DE-AC05-00OR22725, as well as German Research Foundation (DFG) INST 599/1-2.

## References

1. Sekiguchi Y, Kamagata Y, Nakamura K, Ohashi A, Harada H. *Syntrophothermus lipocalidus* gen. nov., sp. nov., a novel thermophilic, syntrophic, fatty-acid-oxidizing anaerobe which utilizes isobutyrate. *Int J Syst Evol Microbiol* 2000; **50**:771-779. [PubMed](#)
2. Sobieraj M, Boone DR. Syntrophomonadaceae. *The Prokaryotes* 2006; **4**:1041-1049. [doi:10.1007/0-387-30744-3\\_37](#)
3. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic



- analysis. *Mol Biol Evol* 2000; **17**:540-552. [PubMed](#)
4. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics* 2002; **18**:452-464. [PubMed](#) [doi:10.1093/bioinformatics/18.3.452](https://doi.org/10.1093/bioinformatics/18.3.452)
5. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol* 2008; **57**:758-771. [PubMed](#) [doi:10.1080/10635150802429642](https://doi.org/10.1080/10635150802429642)
6. Yarza P, Richter M, Peplies J, Euzéby J, Amann R, Schleifer KH, Ludwig W, Glöckner FO, Rosselló-Móra R. The All-Species Living Tree project: A 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst Appl Microbiol* 2008; **31**:241-250. [PubMed](#) [doi:10.1016/j.syapm.2008.07.001](https://doi.org/10.1016/j.syapm.2008.07.001)
7. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How many bootstrap replicates are necessary? *Lect Notes Comput Sci* 2009; **5541**:184-200. [doi:10.1007/978-3-642-02008-7\\_13](https://doi.org/10.1007/978-3-642-02008-7_13)
8. Liolios K, Mavromatis K, Tavernarakis N, Kyrpides NC. The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2008; **36**:D475-D479. [PubMed](#) [doi:10.1093/nar/gkm884](https://doi.org/10.1093/nar/gkm884)
9. Sieber JR, Sims DR, Han C, Kim E, Lykidis A, Lapidus AL, McDonnald E, Rohlin L, Culley DE, Gunsalus R, et al. The genome of *Syntrophomonas wolfei*: new insights into syntrophic metabolism and biohydrogen production. *Environ Microbiol* 2010; **12**:2289-2301.
10. Svetlitsnyi V, Rainey F, Wiegand J. *Thermosyntropho lipolytica* gen. nov., sp. nov., a lipolytic, anaerobic, alkalitolerant, thermophilic bacterium utilizing short- and long-chain fatty acids in syntrophic coculture with a methanogenic archaeum. *Int J Syst Bacteriol* 1996; **46**:1131-1137. [PubMed](#) [doi:10.1099/00207713-46-4-1131](https://doi.org/10.1099/00207713-46-4-1131)
11. Brazelton WJ, Baross JA. Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. *ISME J* 2009; **3**:1420-1424. [PubMed](#) [doi:10.1038/ismej.2009.79](https://doi.org/10.1038/ismej.2009.79)
12. Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A, Takami H, Morita H, Sharma VK, Srivastava TP, et al. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* 2007; **14**:169-181. [PubMed](#) [doi:10.1093/dnares/dsm018](https://doi.org/10.1093/dnares/dsm018)
13. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006; **72**:5069-5072. [PubMed](#) [doi:10.1128/AEM.03006-05](https://doi.org/10.1128/AEM.03006-05)
14. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. [PubMed](#) [doi:10.1038/nbt1360](https://doi.org/10.1038/nbt1360)
15. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. [PubMed](#) [doi:10.1073/pnas.87.12.4576](https://doi.org/10.1073/pnas.87.12.4576)
16. Garrity GM, Holt JG. The Road Map to the Manual. In: Garrity GM, Boone DR, Castenholz RW (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 1. Springer, New York 2001:119-169.
17. Gibbons NE, Murray RGE. Proposals concerning the higher taxa of *Bacteria*. *Int J Syst Bacteriol* 1978; **28**:1-6. [doi:10.1099/00207713-28-1-1](https://doi.org/10.1099/00207713-28-1-1)
18. Validation list 132. List of new names and new combinations previously effectively, but not validly, published. *Int J Syst Evol Microbiol* 2010; **60**:469-472. [doi:10.1099/ijs.0.022855-0](https://doi.org/10.1099/ijs.0.022855-0)
19. Rainey FA. 2009. Class II. Clostridia class nov., p. 736. In P. De Vos, G. Garrity, D. Jones, N. R. Krieg, W. Ludwig, F. A. Rainey, K. H. Schleifer, and W. B. Whitman (ed.), *Bergey's Manual of Systematic Bacteriology*, Second Edition ed, vol. 3. Springer, New York.
20. Prévot AR. in: P. Hauduroy, G. Ehringer, G. Guillot, J. Magrou, A. R. Prévot, Rosset and A. Urbain: *Dictionnaire des Bactéries Pathogènes*, 2nd ed., Masson, Paris, 1953, pp. 1-692.
21. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; **30**:225-420. [doi:10.1099/00207713-30-1-225](https://doi.org/10.1099/00207713-30-1-225)
22. Zhao H, Yang D, Woese CR, Bryant MP. Assignment of fatty acid- $\beta$ -oxidizing syntrophic bacteria to *Syntrophomonadaceae* fam. nov. on the basis of 16S rRNA sequence analyses. *Int J Syst Bacteriol* 1993; **43**:278-286. [PubMed](#) [doi:10.1099/00207713-43-2-278](https://doi.org/10.1099/00207713-43-2-278)

23. Jumas-Bilak E, Roudiere L, Marchandin H. Description of '*Synergistetes*' phyl. nov. and emended description of the phylum '*Deferribacteres*' and of the family *Syntrophomonadaceae*, phylum '*Firmicutes*'. *Int J Syst Evol Microbiol* 2009; **59**:1028-1035. [PubMed](#) [doi:10.1099/ijs.0.006718-0](#)
24. Classification of bacteria and archaea in risk groups. <http://www.baua.de> TRBA 466.
25. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene Ontology: tool for the unification of biology. *Nat Genet* 2000; **25**:25-29. [PubMed](#) [doi:10.1038/75556](#)
26. Chertkov O, Sikorski J, Brambilla E, Lapidus A, Copeland A, Rio TGD, Nolan M, Lucas S, Tice H, Cheng JF, et al. Complete genome sequence of *Aminobacterium colombiense* type strain (ALA-1<sup>T</sup>). *Stand Genomic Sci* 2010; **2**:280-289. [doi:10.4056/signs.902116](#)
27. McInerney MJ, Bryant MP, Hespell RB, Costerton JW. *Syntrophomonas wolfei* gen. nov. sp. nov., an anaerobic, syntrophic, fatty acid-oxidizing bacterium. *Appl Environ Microbiol* 1981; **41**:1029-1039. [PubMed](#)
28. Klenk HP, Göker M. En route to a genome-based classification of Archaea and Bacteria? *Syst Appl Microbiol* 2010; **33**:175-182. [PubMed](#) [doi:10.1016/j.syapm.2010.03.003](#)
29. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, et al. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* 2009; **462**:1056-1060. [PubMed](#) [doi:10.1038/nature08656](#)
30. List of growth media used at DSMZ: [http://www.dsmz.de/microorganisms/media\\_list.php](http://www.dsmz.de/microorganisms/media_list.php).
31. DOE Joint Genome Institute <http://www.jgi.doe.gov>
32. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008; **18**:821-829. [PubMed](#) [doi:10.1101/gr.074492.107](#)
33. Phrap and Phred for Windows, MacOS, Linux, and Unix. <http://www.phrap.com>
34. Sims D, Brettin T, Detter JC, Han C, Lapidus A, Copeland A, Glavina Del Rio T, Nolan M, Chen F, Lucas S, et al. Complete genome sequence of *Kytococcus sedentarius* type strain (541<sup>T</sup>). *Stand Genomic Sci* 2009; **1**:12-20. [doi:10.4056/signs.761](#)
35. Lapidus A, LaButti K, Foster B, Lowry S, Trong S, Goltsman E. POLISHER: An effective tool for using ultra short reads in microbial genome assembly and finishing. AGBT, Marco Island, FL, 2008.
36. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Podigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**:119. [PubMed](#) [doi:10.1186/1471-2105-11-119](#)
37. Pati A, Ivanova N, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: A gene prediction improvement pipeline for microbial genomes. *Nat Methods* 2010; **7**:455-457. [PubMed](#) [doi:10.1038/nmeth.1457](#)
38. Markowitz VM, Ivanova NN, Chen IMA, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; **25**:2271-2278. [PubMed](#) [doi:10.1093/bioinformatics/btp393](#)